View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Columbia University Academic Commons

**SEMI-CONVERGENCE OF AN ITERATIVE ALGORITHM**

**Kathryn N. Vasilaky**

**Abstract.** An iterative method is introduced for solving noisy, ill-conditioned inverse problems. Analysis of the semi-convergence behavior identifies three error components - iteration error, noise error, and initial guess error. A derived expression explains how the three errors are related to each other relative to the number of iterations. The Standard Tikhonov regularization method is just the first iteration of the iterative method and the derived noise damping filter is a generalization of the Standard Tikhonov filter. The derived filter is a function two parameters, a regularization parameter and the iteration number parameter. The new method is tested on image reconstruction from projections simulated data set.

# 1 Introduction

Most of the useful inverse problems are large scale, ill-conditioned linear systems of equations which call for the use of regularization methods and iterative un-regularized methods. Well known classes of iterative un-regularized methods are Algebraic Reconstruction Techniques (ART), [2, 10-13] and Simultaneous Iterative Reconstruction Techniques (SIRT) [ 2 – 8]. The best known regularized methods are Truncated SVD (TSVD) and Tikhonov. In regularized methods a regularization parameter is introduced directly into the model and in un-regularized the iteration number is the regularization parameter. The regularization methods are gaining ground in econometrics and could be useful when the pseudoinverse is needed of a rank deficient matrix, which is a discontinuous mapping of the data [Golub Ed. 4, Section 5.5.5] , as for example, in cluster-robust variance estimation in fixed effects models [16]. In this paper I introduce a method, henceforth referred to as Semi-convergent Tikhonov that employs both types of regularization parameters, a perturbation parameter and an iteration number. The approach can be useful in non-linear least squares fitting in the presence of noisy data, as for example, in parameter identification in the spatial Solow model [18] where the Jacobian may be ill-conditioned.

The new semi-convergent iterative method is proposed, using both types of parameters, for the solution of

$$Ax = b,\ A \in \mathbb{R}^{m\times n},\ m \geq n,\quad b = \bar{b} + e, \qquad (1)$$

where b is contaminated by noise e.

Semi-convergence means that, initially, the iteration vectors approach the noiseless solution to $Ax = \bar{b}$ while continuing iterations lead to the naïve solution contaminated by noise. Semi-convergence is discussed in Natterer, [9, p89].

The advantage of having both types of regularization parameters is that the filter function, which filters out noise, is not as constrained as a filter with only one type of regularization. Of course, now we have the task of choosing two parameters instead of one. I introduce a simple heuristic on how to choose both the regularization and the iteration parameters.

We will compare the proposed method to Truncated SVD (TSVD) and Tikhonov in standard form [1], as well as to Simultaneous Iterative Reconstruction Technique (SIRT), such as Landweber's [2].

## 2 Truncated SVD

Consider the SVD representation for matrix A = $U\Sigma V^T$ , where $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_n, 0 \dots 0) \in \mathbb{R}^{m\times n}$ , with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$, and rank n of A. If $\sigma_{k+1} \dots \sigma_n$ are truncated then $A_k = U\Sigma_k V$ where $\Sigma_k = \mathrm{diag}(\sigma_1, \dots, \sigma_k, 0 \dots 0) \in \mathbb{R}^{m\times n}$ has the condition number $\frac{\sigma_1}{\sigma_k}$ which is smaller than $\frac{\sigma_1}{\sigma_n}$ and the TSVD solution to (1) is:

$$x_k = V\Sigma_k^+ U^T b \text{ where } \Sigma_k^+ = \sigma_1^{-1}, \dots, \sigma_k^{-1}, 0, \dots, 0) \in \mathbb{R}^{m\times n} \quad (2)$$

The i[th] diagonal element of $\Sigma_k^+$ can be written as the i[th diagonal] element of $\Sigma_n^+$ times a filter factor $f_i$, defined as:

$$f_i = \begin{cases} 1 \; for \; \sigma_i \geq \sigma_k \\ 0 \; for \; \sigma_i < \sigma_k \end{cases} \quad (3)$$

The TSVD filter simply cuts off the last n-k components of $\Sigma_n^+$. The TSVD solution to (1) can usually be computed from a Q-R factorization of $A_k$ or directly from expression (2) by computing the SVD of $A_k$ .

## 3 Tikhonov Standard Regularization

Consider the perturbed Least Squares problem:

$$\min_x ||Ax - b||_2^2 + \lambda||x||_2^2, \quad \lambda > 0 \quad (4)$$

The formal solution of (4) is:

$$x_\lambda = A_\lambda b \quad where \; A_\lambda = [A^T A + \lambda I_n]^{-1} A^T. \quad (5)$$

In SVD form, $A_\lambda = V\Sigma_\lambda^+ U^T$, where $\Sigma_\lambda^+ = diag[\frac{1}{\sigma_1}\frac{\sigma_1^2}{\sigma_1^2 + \lambda}, \dots, \frac{1}{\sigma_n}\frac{\sigma_n^2}{\sigma_n^2+\lambda}]$. The i[th] diagonal element of $\Sigma_\lambda^+$ can be written as the i[th] diagonal element of $\Sigma_n^+ = (\sigma_1^{-1}, \dots, \sigma_n^{-1}, 0, \dots, 0) \in \mathbb{R}^{m\times n}$ times the filter factor

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda} \qquad (6)$$

The solution to (4) can be computed by first computing the SVD of A and then $V\Sigma_\lambda^+ U^T$ or by solving the least squares problem

$$\min_x \| \begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \|, \qquad (7)$$

Which is equivalent to (4).

For a more explicit explanation as to how the filters dampen noise due to small $\sigma_i$ let us rewrite the Tikhonov solution (5) as follows:

$$x_1 = V \begin{pmatrix} \frac{\sigma_1^2}{\lambda+\sigma_1^2}\frac{1}{\sigma_1} & \cdots & 0 & & 0 \ldots 0 \\ \vdots & \ddots & & \vdots & \\ 0 & \cdots & \frac{\sigma_n^2}{\lambda+\sigma_n^2}\frac{1}{\sigma_n} & 0 \ldots 0 & \end{pmatrix} U^T b = \sum_{i=1}^{n} \frac{\sigma_i^2}{\lambda+\sigma_i^2} \frac{u_i^T(b^{exact}+e)}{\sigma_i} v_i \qquad (8)$$

Expression (8) exposes how noise is filtered by filter (6).

If we choose $\lambda$ very small then the filter will not dampen the effects of very small $\sigma_i$, since for only $\sigma_i \ll \lambda$ the filter $f_i \approx 0$, but for $\lambda \approx \sigma_i^2$, $f_i \approx 0.5$ so the solution $x_\lambda$ will be noisy if $\sigma_i$ is very small. If we choose $\lambda$ very large then the effects of large $\sigma_i$ will be dampened since for $\sigma_i \gg \lambda$ the filter $f_i \approx 1$, which will distort the original model too much. To put another way, if λ is too large, not enough information in b has been extracted and if λ is too small, only noise is left in the residual.

We can see how the choice of λ affects the Tikhonov solution in the following Figure 1 taken from the Lecture Notes by Per Christian Hansen and reproduced here for convenience:
http://www2.compute.dtu.dk/~pcha/DIP/chap4.pdf

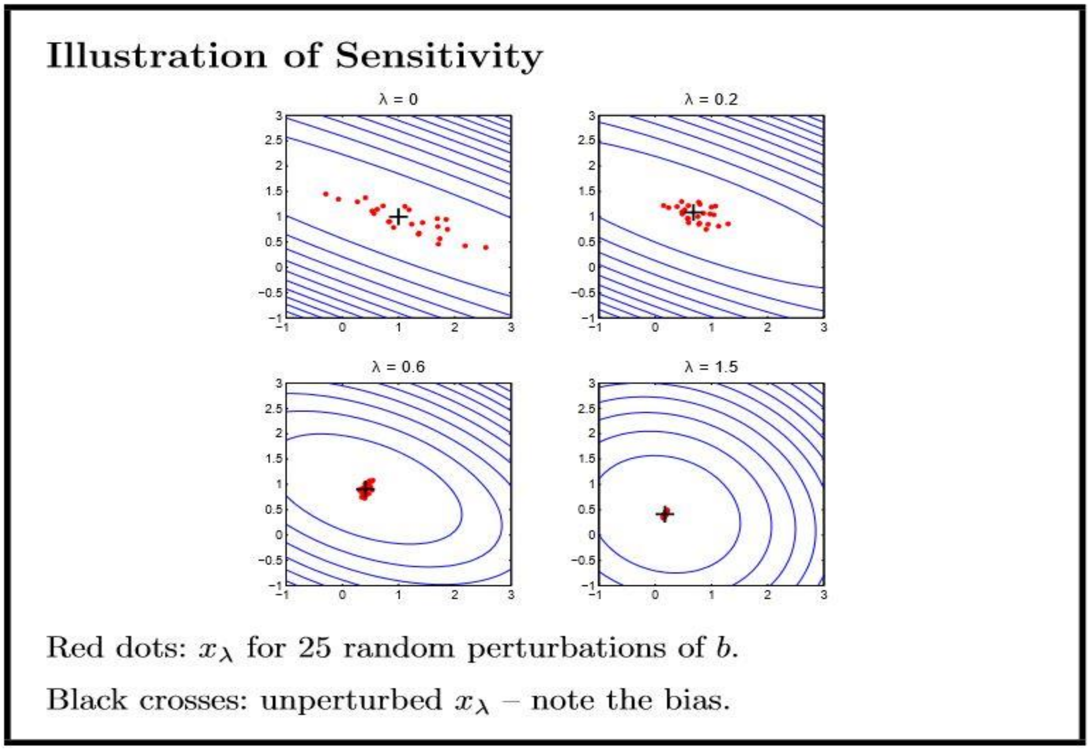



**Figure 1**

Note that in Figure 1 when λ=0, for unperturbed b the solution $x_\lambda$, depicted as +, is (1,1), but for the 25 perturbations of b the solutions (dots) are scattered, see the upper left corner of Figure 1. On the other hand, as λ gets larger, the solution for the unperturbed b shifts toward (0.5 , 0.5), and the solutions (red dots) for perturbed b have stabilized near the wrong solution of (0.5 , 0.5), see in the lower right corner. The same problem holds for TSVD because for each k in TSVD there exists a λ such that $x_k \approx x_\lambda$, C. Hansen [1].

In [1] it is shown that the TSVD step function filter (3) can be seen as an approximation to the smooth regularization filter (6). So choosing the best λ amounts to choosing the best cut off index k in TSVD. Both filters are somewhat crude instruments in damping the solutions to (1). In this paper I introduce a more flexible filter.

## 4 The Semi-Convergent Tikhonov Method

In this paper an iterative method is proposed for the solution of the following Tikhonov regularization problem,

$$\min_x ||Ax - b||_2^2 + \lambda||x - x_0||_2^2, \ \lambda > 0. \tag{9}$$

The iterative method consists of substituting the solution to (9) into $x_0$ to produce the next iterate $x_1$. We then obtain a sequence of vectors $x_0, x_1, \ldots\, x_k \ldots,$ that converges to the noisy solution as $k \to \infty$, and semi-converges to the noiseless solution of $\min_x \|Ax - \bar{b}\|_2^2$, $b = \bar{b} + e$, where e represents noise in the data b.

To solve (9), usually the following two alternative formulations are considered:

$$(A^T A + \lambda I)x = A^T b + \lambda x_0 \tag{10}$$

$$\min_x \left\| \begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix} x - \begin{pmatrix} b \\ \sqrt{\lambda} x_0 \end{pmatrix} \right\|. \tag{11}$$

The first consists of the linear normal equations which are not practical to solve directly. The second shows how to solve (9) stably by treating it as a least squares problem. We will take a new iterative approach. The new iterative method produces a more general filter such that filter (6) is just a special case.

The formal solution to the normal equations (11) is

$$x_\lambda^{[1]} = (A^T A + \lambda I)^{-1} A^T b + \lambda (A^T A + \lambda I)^{-1} x_0 . \tag{12}$$

Substituting $x_1$ $for$ $x_0$ we obtain the second iterate

$$x_\lambda^{[2]} = (A^T A + \lambda I)^{-1} A^T b + \lambda ((A^T A + \lambda I)^{-1})^2 A^T b + \lambda^2 ((A^T A + \lambda I)^{-1})^2 x_0, \tag{13}$$

and so on, the $k^{th}$ iterate is

$$x_\lambda^{[k]} = \sum_{i=1}^{k} \lambda^{i-1} ((A^T A + \lambda I)^{-1})^{i} A^T b + \lambda^k ((A^T A + \lambda I)^{-1})^{k} x_0). \tag{14}$$

**Theorem 1.** Assume $A \in \mathbb{R}^{m\times n}$ is of full rank for m ≥ n, then $x_k$ in (14) converges exponentially to the unique solution of Ax = b for m = n, and to the unique solution of $A^T A x = A^T b$ for m > n.

The proof of Theorem 1 immediately follows from (16) below.

Remark 1: The solution in Theorem 1 is the naïve solution which is sensitive to perturbations of b, hence not very useful when A is severely ill-conditioned. The solution we seek is one that is more stable and one that approximates the noiseless solution of Ax = $\bar{b}$, where b = $\bar{b}$ + e as in (1)

Remark 2: The successive approximations (14) should not be confused with what is known as Iterative Tikhonov regularization:

$x_0 = 0,\ r_0 = b$ , for k = 1, 2,…, $x_k = x_{k-1} + (A^T A + \lambda I)^{-1} A^T r_{k-1},\ r_k = b - Ax_k$ , which, as pointed out in [17], can be considered as a *preconditioned Landweber iteration*.

Substituting the SVD representation for A = $U\Sigma V^T$ in (14), where $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_n, 0 \dots 0) \in \mathbb{R}^{m\times n}$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$, we obtain the following k^th iterate:

$$x_\lambda^{[k]} = V\left(\lambda^{-1}\sum_{i=1}^{k}\begin{pmatrix}(\frac{\lambda}{\lambda+\sigma_1^2})^i & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & (\frac{\lambda}{\lambda+\sigma_n^2})^i\end{pmatrix}\right)V^T A^T b +$$

$$V\begin{pmatrix}(\frac{\lambda}{\lambda+\sigma_1^2})^k & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & (\frac{\lambda}{\lambda+\sigma_n^2})^k\end{pmatrix}V^T x_0 \ . \tag{15}$$

After summing and algebraic manipulation the k^th iterate becomes

$$x_\lambda^{[k]} = V\begin{pmatrix}\frac{1}{\sigma_1}\left(1-(\frac{\lambda}{\lambda+\sigma_1^2})^k\right) & \cdots & 0 & 0 \dots 0\\ \vdots & \ddots & \vdots & \\ 0 & \cdots & \frac{1}{\sigma_n}\left(1-(\frac{\lambda}{\lambda+\sigma_n^2})^k\right) & 0,\dots\ 0\end{pmatrix}U^T b +$$

$$V\begin{pmatrix}(\frac{\lambda}{\lambda+\sigma_1^2})^k & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & (\frac{\lambda}{\lambda+\sigma_n^2})^k\end{pmatrix}V^T x_0. \tag{16}$$

From the first term in (16) we can clearly see that the filter is

$$f_i^{[k]} = 1-\left(\frac{\lambda}{\lambda+\sigma_i^2}\right)^k = \frac{(\lambda+\sigma_i^2)^k - \lambda^k}{(\lambda+\sigma_i^2)^k}. \tag{17}$$

In the standard Tikhonov solution (8) the magnification factor $\frac{1}{\sigma_i}$ was replaced by a gentler factor $\frac{\sigma_i}{\lambda+\sigma_i^2}$ and now with the filter $\frac{(\lambda+\sigma_i^2)^k - \lambda^k}{(\lambda+\sigma_i^2)^k} \approx \frac{k\sigma_i^2}{\lambda+k\sigma_i^2}$ the magnification factor is even gentler $\frac{k\sigma_i}{\lambda+k\sigma_i^2}$. We can do better with an optimal choice of (λ, k) than just with an optimal λ. This is borne out empirically in Figures 6 and 7.

Note the presence of $x_0$ in the solution, so that, in a stochastic setting, it is possible to reformulate this method, under certain distribution assumptions, as a Bayesian method. From (16) we can see that the influence of the initial (a priori) solution $x_0$ diminishes exponentially as k increases.

Remark 3: For k=1 (17) reduces to the Tikhonov filter (6). Thus the Tikhonov method can be viewed as special cases of our iterative method and, by virtue of the results in [1], the TSVD method as well.

Remark 4: (16) proves Theorem 1, since $\frac{1}{\sigma_i}(1-(\frac{\lambda}{\lambda+\sigma_i^2})^k) \to \frac{1}{\sigma_i}, i = 1, \dots, n, \ as\ k \to \infty$ , it follows that

$$x_\lambda^{[k]} \to V\begin{pmatrix} \frac{1}{\sigma_1} & \cdots & 0 & 0 \dots 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n} & 0 \dots 0 \end{pmatrix} U^T b = (A^T A)^{-1} A^T b = A^+ b \text{ as k} \to \infty. \quad (18)$$

The graph of the iterative filter $f_i^{[k]}\ vs.\ \sigma_i$ , for λ= 0.1 and for k = 1, 5, 10, and 20 looks like:

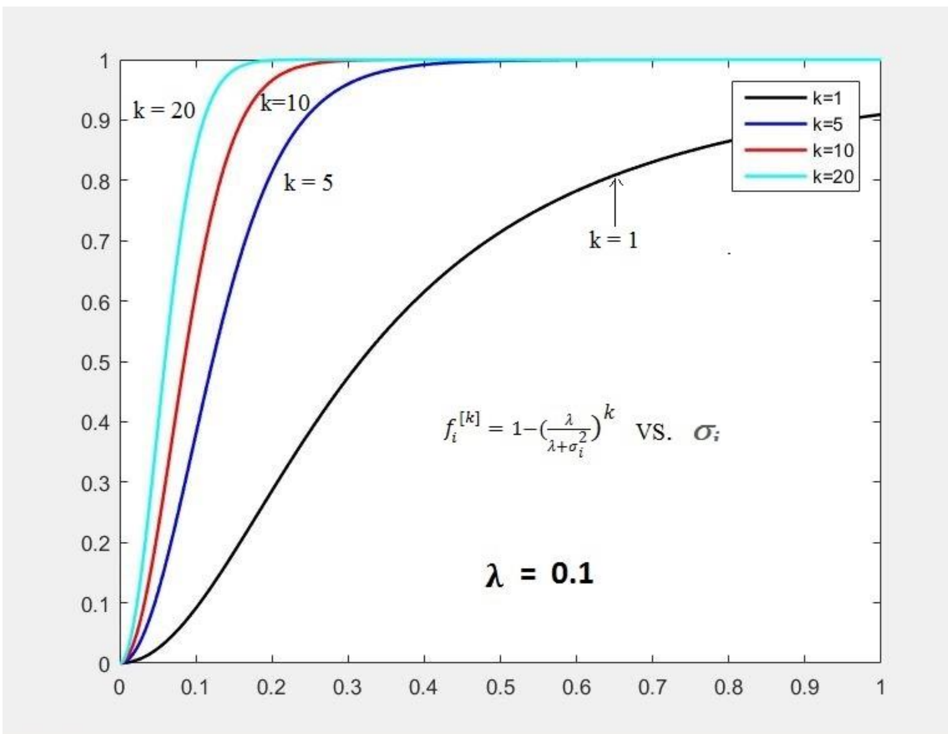



**Figure 3** $f_i^{[k]}\ vs.\ \sigma_i$ ,

The black curve, k=1, is the Tikhonov filter for λ= 0.1

Note that for k=1 (Tikhonov filter) the singular values $\sigma_i = 1$ or larger will be dampened which is undesirable, but for k=20 only singular values less than 0.1 will be dampened.

What is the conclusion? The conclusion is that an optimal choice of the pair (λ , k) in Semi-Convergent Tikhonov will do as well or better that an optimal choice of λ in Standard Tikhonov or an optimal cut off in TSVD. The method can be appropriate in non-linear least squares problems in the presence of noisy data. For example Engbers et al. [18] apply Tikhonov regularization as in (9) in a formulated inverse problem for the identification of production functions from the data in the context of the spatial Solow model. For this non-linear least squares problem, the Jacobian would be A in the normal equations (10), x would be the increment vector of parameter values and b would be the increment residual vector. The difficulty of choosing the regularization parameter λ, can circumvented by the Semi-Convergent Tikhonov, since the choice of λ is not worth optimizing because the number of iterations adjusts to the chosen parameter λ using the heuristic in Section 6.

## 5 Error Analysis

Let $\bar{x}$ denote the minimum norm solution with noise-free right hand side $\min_x \left\|Ax - \bar{b}\right\|_2^2$ , where b = $\bar{b} + e$ , and e is the noise component of b. It is straightforward to show that

$$\bar{x} = V \begin{pmatrix} \frac{1}{\sigma_1} & \cdots & 0\;0 \ldots 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n}\,0 \ldots 0 \end{pmatrix} U^T \bar{b}. \qquad (19)$$

Hence the difference between the k<sup>th</sup> iterate and the noiseless solution to the unperturbed problem is:

$$\mathrm{f}(\lambda, \mathrm{k}) \equiv x_\lambda^{[k]} - \bar{x} = -V \begin{pmatrix} \frac{1}{\sigma_1}\left(\frac{\lambda}{\lambda+\sigma_1^2}\right)^k & \cdots & 0 & 0 \ldots 0 \\ \vdots & \ddots & \vdots & \\ 0 & \cdots & \frac{1}{\sigma_n}\left(\frac{\lambda}{\lambda+\sigma_n^2}\right)^k & 0 \ldots 0 \end{pmatrix} U^T \bar{b} +$$

$$V \begin{pmatrix} \frac{1}{\sigma_1} - \frac{1}{\sigma_1}\left(\frac{\lambda}{\lambda+\sigma_1^2}\right)^k & \cdots & 0 & 0 \ldots 0 \\ \vdots & \ddots & \vdots & \\ 0 & \cdots & \frac{1}{\sigma_n} - \frac{1}{\sigma_n}\left(\frac{\lambda}{\lambda+\sigma_n^2}\right)^k & 0 \ldots 0 \end{pmatrix} U^T e +$$

$$V \begin{pmatrix} \left(\frac{\lambda}{\lambda+\sigma_1^2}\right)^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \left(\frac{\lambda}{\lambda+\sigma_n^2}\right)^k \end{pmatrix} V^T x_0 \,. \qquad (20)$$

The first term is $\bar{x}_{\lambda}^{[k]} - \bar{x}$, the iteration error, and the second term is $x_{\lambda}^{[k]} - \bar{x}_{\lambda}^{[k]}$, the noise-error, and the third term is the initial solution error. Other names are used in the literature for the first two errors. For example, "approximation error" and "data error". It is clear from (20) that the error norm, $||x_{\lambda}^{[k]} - \bar{x}||_2$, decreases initially where the regularization error $\bar{x}_{\lambda}^{[k]} - \bar{x}$ dominates, whereas the error norm starts to increase after a certain stage and the noise error, $x_{\lambda}^{[k]} - \bar{x}_{\lambda}^{[k]}$, starts to dominate. The goal is to choose λ and k such that the norm of the sum of the three errors is an absolute minimum. It is clear that for any closed and bounded domain, D: $0 \le \lambda \le \lambda^*$, $1 \le k \le k^*$, there exists a point $(\hat{\lambda}, \hat{k})$ in D such that $||f(\hat{\lambda}, \hat{k})||_2$ is an absolute minimum in D. Since $||f(\lambda, k)||_2$ is clearly differentiable with respect to λ and k and D is closed and bounded, that guarantees that $||f(\lambda, k)||_2$ attains an absolute global minimum in D.

To derive an explicit closed form solution for $(\hat{\lambda}, \hat{k})$ from $||f(\lambda, k)||_2$ in terms of $||e||_2$ does not appear to be feasible and to obtain a numerical solution would require knowledge of the error term e, which is usually not known in practice. However by virtue of the fact that the Standard Tikhonov is a special case of Semi-Convergent Tikhonov, it follows that $||f(\hat{\lambda}, \hat{k})||_2 \le ||f(\lambda^{opt}, 1)||_2$ where $\lambda^{opt}$ is the optimal λ in the Standard Tikhonov method. Recall that $f(\lambda, 1)$ is the error for the Standard Tikhonov, which means that the iterative method is as good as or better than the Standard Tikhonov method. In section 7 a heuristic method is proposed for choosing λ and a stopping iteration number k.

## 6 How to Choose Parameter λ and Number of Iterations k.

We will make use of the relative residual error $\frac{||Ax_{\lambda}^{[k]} - b||_2}{||b||_2}$ to choose λand k, fortunately, we do not need to compute the iterative solution $x_{\lambda}^{[k]}$ to compute the residual error. Since

$$Ax_{\lambda}^{[k]} - (\bar{b} + e) = Ax_{\lambda}^{[k]} - A\bar{x} - e = A(x_{\lambda}^{[k]} - \bar{x}) + e \qquad (25)$$

substituting $U\Sigma V^T$ for A in (25) and (20) for $x_{\lambda}^{[k]} - \bar{x}$, with $x_0 = 0$, after some algebraic manipulation, we obtain

$$Ax_{\lambda}^{[k]} - b = -U\begin{pmatrix} (\frac{\lambda}{\lambda+\sigma_1^2})^k & \cdots & 0 \ldots \ldots .0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (\frac{\lambda}{\lambda+\sigma_n^2})^k \ldots 0 \end{pmatrix} U^T b$$

Since U is unitary, we can drop the U on the left to obtain the residual error:

$$\mathrm{RE}(\lambda,k) \equiv \left|\left| \begin{pmatrix} (\frac{\lambda}{\lambda+\sigma_1^2})^k & \cdots & 0 \ldots \ldots .0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (\frac{\lambda}{\lambda+\sigma_n^2})^k \ldots 0 \end{pmatrix} U^T b \right|\right| \tag{26}$$

The heuristic for choosing λ and k is to compute the relative residual error RE(λ,k) using (26) for several values of λ in an iteration loop running some reasonable number times, say 50 and plotting in the same graph the curves RE($\lambda_1$, 100) $vs.\ k$, …, RE($\lambda_p$, 100) $vs.\ k$ ; we then choose the smallest λ for which the curve exhibits an L-shape curve, see Figure 5 below. The choice of k, is decided by where the curve has maximum curvature, what I refer to as the elbow of the curve. Once we have chosen λ and k the matrices needed to compute $x_\lambda^{[k]}$ by (16) are already in memory. The maximum curvature point is simple to compute and the algorithm for computing $x_\lambda^{[k]}$ is only a few lines of code.

My Python code for the heuristic and computation of $x_\lambda^{[k]}$ is available at Python Software Foundation, Inverse Problem, https://pypi.python.org/pypi/InverseProblem/1.0 .

# 7 Un-regularized Iterative Method

(ART and SIRT Methods are the best known un-regularized methods. We will consider a SIRT method. In particular we will derive the Landweber filter and compare it to our filter. The Landweber method generates the following sequence:

$$\mathrm{x_k} = \mathrm{x_{k-1}} + \lambda A^T(b - \mathrm{A}\mathrm{x_{k-1}})\ ;\ \mathrm{x_0} = 0 \quad \text{k=1, 2…} \tag{21}$$

where $\lambda$ is a fixed relaxation parameter. After k iterations we have

$$\mathrm{x_k} = (\mathbf{I} - \lambda A^T A)^k \mathrm{x_0} + \sum_{i=0}^{k-1} (\mathbf{I} - \lambda A^T A)^i (A^T b)$$

$$= (A^T A)^{-1}(\mathbf{I} - (\mathbf{I} - \lambda A^T A)^k)(A^T b) \tag{22}$$

Where we summed the geometric series in $(\mathbf{I} - \lambda A^T A)^i$ to obtain (22). Since $A^T A = V(\Sigma^T \Sigma)V^T$ and I = V$V^T$ then it follows

$$\mathrm{x_k} = \sum_{i=1}^{k} (1 - \left(1 - \lambda\sigma_i^2\right)^k)\,(\frac{u_i^T b}{\sigma_i}) v_i \tag{23}$$

From (23) we see that the Landweber filter factors are

$$1 - \left(1 - \lambda\sigma_i^2\right)^k . \tag{24}$$

In order for $|1 - \lambda\sigma_i^2| < 1$ requires $0 < \lambda\sigma_i^2 < 2$ which implies $\lambda < \frac{2}{\sigma_i^2}$ for all i, and in particular $\lambda < \frac{2}{\sigma_1^2}$. If $\sigma_1$ is very large then the relaxation parameter $\lambda$ is very small which slows down convergence. Note that 1 - $\left(1 - \lambda\sigma_i^2\right)^k \approx k\lambda\sigma_i^2$ which means $\lambda \to 0$ as k -> ∞, it is clear that the effects of $\lambda$ and k are equivalent. On the other hand, $\frac{(\lambda+\sigma_i^2)^k - \lambda^k}{(\lambda+\sigma_i^2)^k} \approx \frac{k\sigma_i^2}{\lambda+k\sigma_i^2}$ -> 1 as k -> ∞ for any fixed $\lambda$, hence $\lambda$ is not forced to go to 0 as k->∞ . Also the only way the Tikhonov filter $\frac{\sigma_i^2}{\lambda+\sigma_i^2}$ -> 1 for any $\sigma_i^2$ is if $\lambda \to 0$ . But forcing $\lambda$ to go to zero means that the solution will become noisy. That is the reason the Semi-Convergent Tikhonov filter is a more flexible than the other filters under discussion. Yet another attractive feature of the Semi-Convergent Tikhonov is its simplicity, the heuristic and the iterative solution require just a few line of code. The unattractive feature the new method is the computation of SVD which has complexity O(m$n^2$). Fortunately the computational expense is becoming less and less of a concern, for algorithms of order O(m$n^2$). Also the computation of SVD is needed only once for both the heuristic and the iterative solution. A change of $\lambda$ or data b and as well as each iteration can reuse the stored SVD. Having stored the SVD decomposition, each subsequent computation require 2mn flops. Of course, if A is sparse, the iterative approach is computationally competitive with the other methods.

# 8 Experimental Results

**Image Tomography.** We now consider a test problem from the field of tomographic image reconstruction from projections, such as CT scans. The motive for the choice of this test problem is convenience, since the simulated data set is readily available from a MatLab package. Also because the projections are in random directions and not uniformly spaced as in commercial CT scans. Random projections occur more often in nature as for instance in Seismology, Volcanology, and Econometrics. Per Christian Hansen's MatLab package provides the function [A,b,x] = tomo(N,f) which creates a simple two-dimensional tomography test problem. A 2D domain [0,N] x [0,N] is divided into N^2 cells of unit size, and a total of round(f*N^2) rays in random directions penetrate this domain. The default value is f = 1. Each entry in matrix A= ( $a_{ij}$) denotes the calculated value of the length of the intersection of ray i with cell j. Each cell is assigned a value stored in the vector x.

For each ray, the corresponding element in the right-hand side b is the approximate discrete line integral along the $i^{th}$ ray, the $i^{th}$ component of b is $b_i = \sum_{j=1}^{N^2} a_{ij} x_j$. The discrete line integrals are also known as the discrete Radon transforms. Hence the problem becomes A$x$ = b, and the solution x is known. The original image can be visualized in Mat Lab by an algorithm, reshape(x, N, N) or in Python imshow (x.reshape(N,N)). For our test, we perturb the right-hand-side b by Gaussian noise with zero mean and standard deviation 0.1. We choose N=64, of course a larger N results in a better resolution, but requires more memory and computation. Hence the number of pixels is 64*64 = 4096 and the dimensions of A, b, x are 4096-by-4096, 4096-by-1, and 4096-by1, respectively.

Figure 5 below shows the Relative Residual Errors for several values of λ in this Tomography Example.

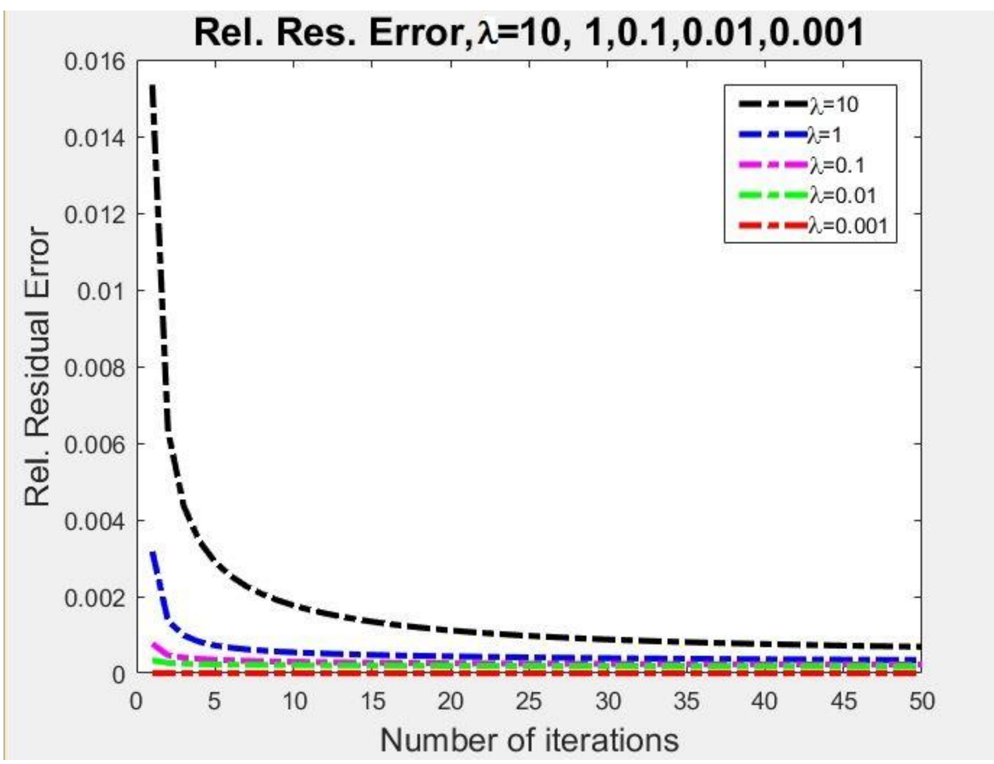
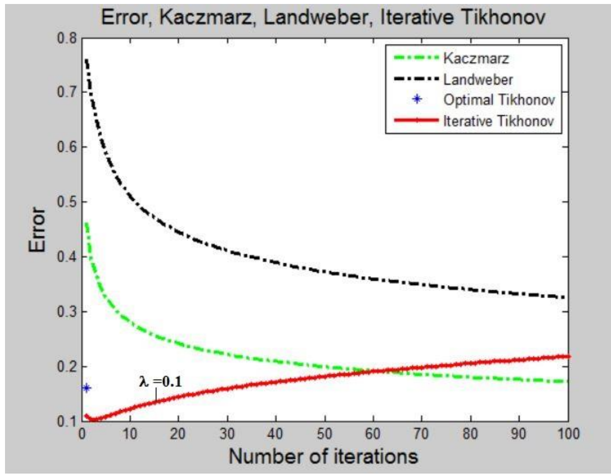



**Figure 5 Iterations vs. relative residual error** $|| Ax_{\lambda}^{[k]} - \mathbf{b}||_2 / || \mathbf{b}||_2$

Figure 5 displays the relative residual errors vs. k for the reconstruction algorithm of an image from random projections in the tomography test example above. Our choice of λ would be 0.1 (red curve) the smallest λ for which we have clear L-shape curve and the choice of k would be about 3 at the elbow of the red curve. We can also choose λ =1 (blue curve) where k increases to about 5 or 6. There is no need to have an optimal λ, as in Standard Tikhonov, because the number if iterations increases with λ so that the choice of λ determines the optimal number of iterations. Nevertheless, the smallest λ or near it that produces a concave L-shape curve is preferred.

For comparison, we display below the relative error curves $||x_{\lambda}^{[k]} - \bar{x}||_2/||\bar{x}||_2$ vs. the number iterations, k, for the Landweber, Kaczmarz, optimal λ Standard Tikhonov and the Semi-Convergent Tikhonov, λ=0.1, λ= 0.5. Since Standard Tikhonov is only one iteration we display one point using an optimal λ based on the L-curve method. Note that for λ = 0.1 the error is minimum at k=4 and for λ = 0.5 the k moved up to k=6 as predicted by our heuristic in Figure 5 above.



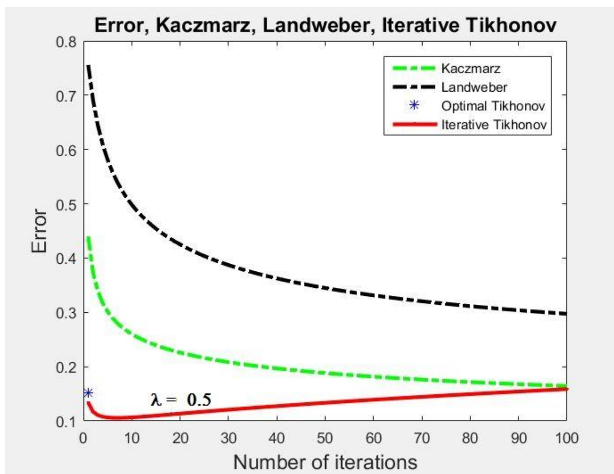
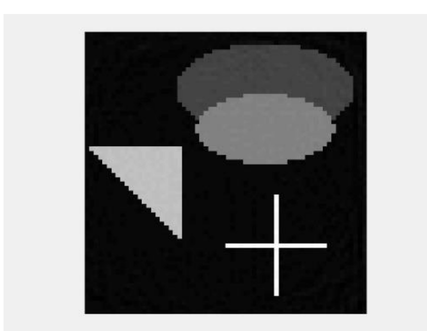




**Figure** 6, $||x_{\lambda}^{[k]} - \bar{x}||_2/||\bar{x}||_2$ vs, k; λ=0.1 **Figure 7,** $||x_{\lambda}^{[k]} - \bar{x}||_2/||\bar{x}||_2$ vs, k; λ=0.5

The following figures shows the original image and the reconstructed images by means of the MatLab back-slash operator, the l_curve( ) and tikhonov( ) algorithms from Per Christian Hansen's MatLab package, as well as the Kaczmarz and Landweber algorithms, in the same package. The right-hand-side b has been perturbed by Gaussian random noise with mean 0 and standard deviation 0.1.



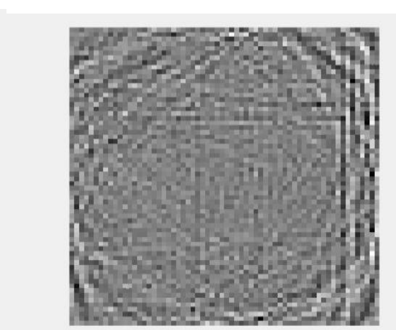
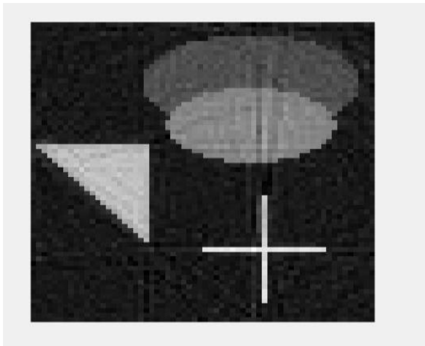


**Figure 8. Noiseless** **Figure 9. MatLab back-slash solution**



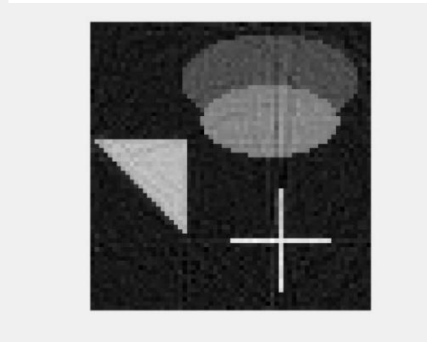
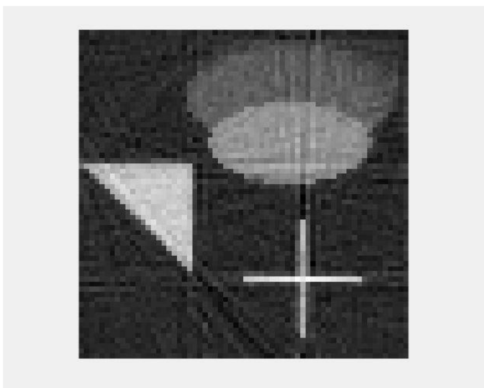


**Figure 10. L-curve() opt.– tikhonov()** **Figure 11. Kaczmarz, k=200**



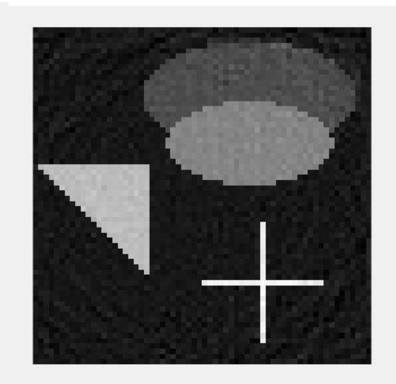


**Figure 12. Landweber, k=200** **Figure 13. Iterative λ=0.1, k=4**

Can the Semi-Convergent Tikhonov improve on the Standard Tikhonov or TSVD? The answer is yes. If we take the Standard Tikhonov optimal solution in Figure 10 and use it as the initial solution $x_0$ in Semi-Convergent Tikhonov we obtain Figure 13, after 3 iterations. A visible improvement.

# 9 Conclusion

The introduced Semi-Convergent Tikhonov method for regularization of problem (1) generates a sequence that semi- converges to the unperturbed noiseless solution and converges to the naïve solution. The Standard Tikhonov regularization is just a special case of Semi-Convergent Tikhonov. By implication, Semi-Convergent Tikhonov can be viewed as a generalization of TSVD when matrix A has a well-determined numerical rank. In addition, since the regularization parameter of ART and SIRT methods is the stopping number, the Semi-Convergent Tikhonov can also be viewed as an iterative method for semi-convergence. The introduced method is suitable if the initial guess may not be reliable hence the regularization parameter would need to be very small, resulting in a noisy solution. The semi-convergent Tikhonov circumvents this problem since the influence of the guess goes to zero exponentially and the regularization parameter does not have to be small. The computed SVD is utilized in both, the heurist for choosing both parameters and the computation of the semi-convergent sequence. If matrix A is fixed but the data and/or regularization parameter changes there is no need to compute SVD again. The heuristic and each iteration require approximately 2mn flops. Simulation experiments support the conjecture that the two parameter method has better error reduction capability than single parameter methods.

## References


[1] Per Christian Hansen, The Truncated SVD as a Method for regularization, *BIT 27 (1987), 534-553*. http://infolab.stanford.edu/pub/cstr/reports/na/m/86/36/NA-M-86-36.pdf

[2] Per Christian Hansen, Maria Saxild-Hansen, AIR Tools – A MATLAB package of algebraic iterative reconstruction methods, *Journal of Computational and Applied Mathematics 236 (2012) 2167-2178*

[3] Elfving Tommy., Nikazad Touraj., Hansen Per Christian. Semi-Convergence and Relaxation Parameters for a Class Of SIRT Algorithms. *Electronic Transactions on Numerical Analysis. Volume 37, pp. 321-336, 2010.*

[4] Bertaro M. and Boccacci P., Introduction to Inverse Problems in Imaging, *Institute of Physics Publishing*, Bristol, 1998.

[5] Engl H. W., Hanke M., and Neubauer A., Regularization of Inverse Problems, *Kluwer, Dordrecht*, The Netherlands, 1996.

[6] Hansen P. C., Rank-Deficient and Discrete Ill-Posed Problems, *SIAM*, Philadelphia, 1998.

[7] Piana M. and Bertero M., Projected Landweber method and preconditioning, *Inverse Problems, 13 (1997), pp. 441–464.*

[8] Van Der Sluis A. and D H. A. Van Der Vorst D H. A., SIRT- and CG-type methods for the iterative solution of sparse linear least-squares problems, *Linear Algebra Appl., 130 (1990), pp. 257–303.*

[9] Natterer, F. *The Mathematics of Computerized Tomography*, John Wiley, New York, 1986.

[10] Gordon R., Bender R., and Herman G. T., *Algebraic reconstruction techniques (ART) for three dimensional electron microscopy and X-ray photography.*

[11] Gordon R. and Herman G. T., "Reconstruction of pictures from their projections," Commun. Assoc. Comput. Mach., vol. 14, pp. 759-768, 1971.

[12] Herman G. T. and Rowland S., "Resolution in ART: An experimental investigation of the resolving power of an algebraic picture reconstruction," *J. Theor. Biol., vol. 33, pp. 213-233, 1971.*

[13] Herman G. T., Lent A., and Rowland S., "ART: Mathematics and applications: A report on the mathematical foundations and on applicability to real data of the algebraic reconstruction techniques," *J. Theor. Biol., vol. 43, pp. l-32, 1973.*

[14] Hansen Per Christian, Regularization Tools, A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems, Version 4.0 for Matlab 7.3, http://www.mathworks.com/matlabcentral/fileexchange

[15] Gene H. Golub and Charles F. Van Loan, Matrix Computation 3rd edition, *The John Hopkins University Press.*

[16] James E. Pustejovsky and Elizabeth Tipton, "Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models", Working Paper, Department of Educational Psychology, University of Texas at Austin and Department of Human Development, Teachers College, Columbia University. http://arxiv.org/pdf/1601.01981.pdf

[17] Arnold Neumaier, Solving Ill-Conditioned and Singular Linear Systems: A Tutorial on Regularization, *SIAM Rev., 40(3), 636–666*. http://www.mat.univie.ac.at/~neum/ms/regtutorial.pdf

[18] Ralf Engbers, Martin Burger and Vincenzo Capasso, Inverse problems in geographical economics: parameter identification in the spatial Solow model, *Philosophical Transactions of the Royal Society, February 15, 2016.*